AD-A113 877 NEW MEXICO UNIV ALBUQUERQUE F/G 12/1
SMALL SAMPLE PROPERTIES OF ROBUST ANALYSIS OF VARIANCE FOR FIXE--ETC(U)
APR 82 R M SCHRADER, J W MCKEAN DAAG29-79-C-0022

UNCLASSIFIED ARO-16064.4-M NL

1 OF 1
AD A
113877

END
DATE
FILMED
05-82
DTIC

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(11)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 16064.4-M	2. GOVT ACCESSION NO. AD-A113 877	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Small Sample Properties of Robust Analysis of Variance for Fixed and Random Effects Models		5. TYPE OF REPORT & PERIOD COVERED Final: 1 May 79 - 30 Apr 81
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Ronald M. Schrader, Joseph W. McKean		8. CONTRACT OR GRANT NUMBER(s) DAAG29 79 C 0022
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of New Mexico, Albuquerque, NM 87131		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office, Post Office Box 12211, Research Triangle Park, NC 27709		12. REPORT DATE Apr 82
		13. NUMBER OF PAGES 6
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC ELECTE APR 26 1982 D

H

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

robust analysis
random effect models
statistical linear models
least squares methodology

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The major shortcoming of robust methodology in statistical linear models has been a limitation primarily to parameter estimation in fixed effects models. The major competitor, classical least squares methodology, by contrast offers a unified treatment of estimation, testing, and multiple comparisons techniques in a wide range of fixed and random effects models. Research efforts have been directed toward filling the void in robust methods in order to provide a complete alternative to least squares procedures. This report summarizes the results.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A113877

DTIC FILE COPY

82 04 26 059

FINAL REPORT

1. ARO Proposal Number: 16064-M

2. Title of Proposal: Small Sample Properties of Robust Analysis of Variance for Fixed and Random Effects Models.

3. Contract or Grant Number: DAAG 29-79-C-0022

4. Name of Institution: University of New Mexico

5. Authors of Report: Ronald M. Schrader and Joseph W. McKean

6. List of Manuscripts Submitted or Published under ARO Sponsorship:

1. Schrader, R.M. and Hettmansperger, T.P. (1980). Robust analysis of variance based upon a likelihood ratio criterion. Biometrika 67, 93-101.

2. McKean, J.W. and Schrader, R.M. (1980). The geometry of robust procedures in linear models. Jour. Roy. Statist. Soc. B, 42, 366-371.

3. McKean, J.W. and Schrader, R.M. (1981). A comparison of methods for studentizing the sample median. Submitted to Jour. Roy. Statist. Soc. B.

4. Schrader, R.M. and McKean, J.W. (1981). Least absolute errors analysis of variance. Submitted to Jour. Amer. Statist. Assoc.

5. McKean, J.W. and Schrader, R.M. (1981). The use and interpretation of robust analysis of variance. Proceedings of ARO Conference on Modern Data Analysis, New York: Academic Press.

6. Hettmansperger, T.P. and McKean, J.W. (1981). A geometric interpretation of inferences based on ranks in the linear model. Submitted to Jour. Amer. Statist. Assoc.

7. McKean, J.W. and Schrader, R.M. (1981). An efficient and stable algorithm for robust estimation and analysis of variance. Submitted to SIAM Jour. Num. Anal.

7. Scientific Personnel Supported by This Project and Degrees Awarded:

Western Michigan University:

- Joseph W. McKean
- James B. Davis (Masters Degree)
- Mark Gerger
- Richard Pretty
- S. Sreyams

University of New Mexico:

- Ronald M. Schrader
- Rebecca (Cote) Gore (Masters Degree)

8. <u>The Problem Studied</u>:

The major shortcoming of robust methodology in statistical linear models has been a limitation primarily to parameter estimation in fixed effects models. The major competitior, classical least squares methodology, by contrast offers a unified treatment of estimation, testing, and multiple comparisons techniques in a wide range of fixed and random effects models. Our research efforts have been directed toward filling the void in robust methods in order to provide a complete alternative to least squares procedures.

Define the linear model

$$Y = X\beta + \varepsilon \tag{1}$$

where Y is an $n\times 1$ vector of observations, X is an $n\times p$ design matrix of rank k , β is a $p\times 1$ vector of unknown parameters, and ε is an $n\times 1$ vector of random errors. We will consider at present the fixed effects case where the components of ε are independent; the random effects case allows a more general covariance structure and will be considered later. A comprehensive review and discussion appears in McKean and Schrader (1981).

A least squares estimate of β , $\hat{\beta}_{LS}$, is found by minimizing the dispersion function

$$D_{LS}(\beta) = \sum(y_i - x_i\beta)^2 \tag{2}$$

where the x_i is the i^{th} row of X . An M-estimate, $\hat{\beta}_M$, is found by minimizing

$$D_M(\beta) = \sum\rho(y_i - x_i\beta) \tag{3}$$

for an appropriate function ρ , see Huber (1973). An R-estimate, $\hat{\beta}_R$, is found by minimizing

$$D_R(\beta) = \sum a(R|y_i - x_i\beta|)(y_i - x_i\beta) \tag{4}$$

where $a(\cdot)$ is an appropriate score function on the integers and $R|u_i|$ denotes the rank of $|u_i|$ among $|u_1|,\ldots,|u_n|$; see Jaeckel (1972). Note that an ℓ_1 estimate corresponds either to $a(i) \equiv 1$ or $\rho(x) = |x|$ in (4) or (3), respectively; hence an ℓ_1 estimate is both an R- and M-estimate. Under regularity conditions, all of these estimators are asymptotically normal with variance-covariance matrix $K^2(X'X)^-$ where K is a scale parameter which depends upon the distribution of errors ε_i and the estimate selected, and $(X'X)^-$ is an appropriate generalized inverse of $(X'X)$.

General linear hypotheses are expressed as

$$H_0 : H\beta = 0 \text{ vs. } H_1 : H\beta \neq 0 \quad (5)$$

for a $q \times p$ matrix of rank q. The model (1) is the full model and (1) subject to H_0 is the reduced model. The classical F-statistic for testing H_0 is

$$F_{LS} = [(D_{LS}(R) - D_{LS}(F))/(k-q)]/\hat{\sigma}^2 \quad (6)$$

where $D_{LS}(R)$ and $D_{LS}(F)$ denote minimum vlaues of (2) under reduced and full models, respectively, and $\hat{\sigma}^2$ is an estimate of $\mathrm{Var}(\varepsilon_i)$. We have proposed robust test statistics, F_M and F_R, similar to F_{LS} where

$$F_M = [(D_M(R) - D_M(F))/(k-q)]/\hat{\lambda}_M \quad (7)$$

with $D_M(R)$ and $D_M(F)$ the reduced and full model values of (3). F_R is defined similarly.

We published asymptotic distribution theory of F_R and F_M in McKean and Hettmansperger (1976) and Schrader and Hettmansperger (1980). This asymptotic theory, along with asymptotic theory of the estimator, $\hat{\beta}$, provides useful guidelines on appropriate standardizing constants and null distribution theory. We investigated small sample properties of these procedures during the contract. Specifically, various proposals for estimating the standardizing constants K and λ were examined. The intent of the research was to provide a good approximation to the small sample distributions of $\hat{\beta}$, F_R, and F_M in order to provide reasonable confidence and inference procedures.

9. <u>Summary of Major Results</u>:

Asymptotic theory for F_R was developed previously by McKean and Hettmansperger (1976,1978). During the term of this contract, Schrader and Hettmansperger (1980) established similar asymptotic theory for F_M and certain variants of F_M. In this work the connection between F_M and $\hat{\beta}_M$ was shown to be the same as the connection between actual maximum likelihood estimates and likelihood ratio tests.

The ℓ_1, or least absolute errors, estimate has received considerable attention in recent years. As noted previously this is technically both an M- and an R-estimate. Because the score function involved fails to meet certain smoothness criteria, the distribution theory cited above for F_M and F_R does not apply to a similar "F-ratio." In Schrader and McKean (1981) we established asymptotic theory for the ℓ_1 procedure and investigated its small sample behavior.

A coordinate-free approach to classical linear models theory, as discussed by Kruskal (1968) greatly enhances the interpretability of least squares techniques.

In McKean and Schrader (1980a), we developed a similar geometric formulation of robust methods, providing a corresponding simple interpretation. In this work we discussed ideas of estimable functions and testable hypotheses in cases where expectations of errors need not exist, and demonstrated that robust methods involve only replacing the ℓ_2 norm by another appropriate norm.

In order to gain insight into the problem of estimating the standardizing constants K and λ, we began Monte Carlo work with the ℓ_1 estimate, which is both an R- and M-estimate. The constant λ in this case is $\lambda = [4f(\eta)]^{-1}$, and the constant K is $[2f(\eta)]^{-1}$ where f is the underlying density of the errors and η is the median of this distribution. Since $K = \frac{1}{2}\lambda$, as is the case for all R- and M-estimates, we began with the simple location problem and examined how well various estimates of K "Studentize" the sample median (the ℓ_1 estimate in this case). In a large Monte Carlo study of this problem McKean and Schrader (1980b) we found that two estimates of K, hence also of λ, served to standardize the median in a much more stable manner than several others; the Bootstrap estimate of Efron (1979) and a standardized confidence interval length similar to a proposal by Lehmann (1963). Surprisingly, we discovered that the best small sample approximating distribution for a Studentized median is the standard normal; it would be expected that a more heavy tailed distribution such as Student's t would be appropriate.

We proceeded to an in depth study of the robust F-test for ℓ_1 estimates in Schrader and McKean (1981). The estimates of λ which performed well in the study of the median were employed in this study. We also examined a large number of both numerator and demoninator degrees of freedom, both of which should exert considerable influence on the behavior of the procedure. In this study we discovered that both quadratic forms in ℓ_1 estimates and F_{LS} can be so unstable for certain error distributions that they are almost impossible to standardize adequately. The ratio based upon reduction in absolute errors behaved in a reasonably stable manner, by contrast. Also surprising for ℓ_1, as it was for the median, was that the asymptotic distribution (chi-squared) is the appropriate standardizing distribution; again, one would expect a more direct influence of denominator degrees of freedom.

We developed efficient and stable algorithms to perform robust analysis for general non-full rank linear models; see McKean and Schrader (1981). Extensive use was made of the state-of-the-art numerical software package LINPACK (Dongerra, et.al., 1979).

In a comparison between classical outlier detection methods (John, 1978) and robust methodology, we demonstrated that robust methods perform automatically the desired outlier detection and inference (McKean and Schrader, 1981). Classical methods are much more complicated and subjective, and are less reliable.

The statistic F_R is only one of several proposals for an R-estimate based analysis of variance. Hettmansperger and McKean (1981) presented a unified geometric approach to various of these methods. In a Monte Carlo study they demonstrated that F_R behaved most stably across error distributions and design configurations.

10. References:

1. Dongerra, J.J., Bunch, J.R., Moler, C.B. and Stewart, G.W. (1979). LINPACK Users Guide, Philadelphia: Society for Industrial and Applied Mathematics.

2. Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. Ann. Statist. 7, 1-26.

3. Hettmansperger, T.P. and McKean, J.W. (1981). A geometric interpretation of inferences based on ranks in the linear model. Submitted to Jour. Amer. Statist. Assoc.

4. Huber, P.J. (1973). Robust regression: Asymptotics, conjectures, and Monte Carlo. Ann. Statist. 1, 799-821.

5. Jaeckel, L.A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. Ann. Math. Statist. 43, 1449-58.

6. John, J.A. (1978). Outliers in factorial experiments. Appl. Statist. 27, 111-119.

7. Kruskal, W. (1968). When are Gauss-Markov and least squares estimators identical? Ann. Math. Statist. 39, 70-75.

8. Lehmann, E.L. (1963). Nonparametric confidence intervals for a shift parameter. Ann. Math. Statist. 34, 1507-12.

9. McKean, J.W. and Hettmansperger, T.P. (1976). Tests of hypothesis based on ranks in the general linear model. Commun. Statist. A., 693-709.

10. McKean, J.W. and Hettmansperger, T.P. (1978). A robust analysis of the general linear model based on one step R-estimates. Biometrika 65, 571-79.

11. McKean, J.W. and Schrader, R.M. (1980a). The geometry of robust procedures in linear models. Jour. Roy. Statist. Soc. B. 42, 366-71.

12. McKean, J.W. and Schrader, R.M. (1980b). A comparison of methods for studentizing the sample median. Submitted to Jour. Roy. Statist. Soc. B.

13. McKean, J.W. and Schrader, R.M. (1981). The use and interpretation of robust analysis of variance. Proceedings of ARO Conference on Modern Data Analysis. New York: Academic Press.

14. Schrader, R.M. and Hettmansperger, T.P. (1980). Robust analysis of variance based upon a likelihood ratio criterion. Biometrika 67, 93-101.

15. Schrader, R.M. and McKean, J.W. (1981). Least absolute errors analysis of variance. Submitted to Jour. Amer. Statist. Assoc.